

## Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode *Gaussian naïve bayes*

Syamsul Mujahidin<sup>1\*</sup>, Bagus Prasetyo<sup>2</sup>, Muchammad Chandra Cahyo Utomo<sup>3</sup>

<sup>1,2,3</sup>Program Studi Informatika, Jurusan Matematika dan Teknologi Informasi, Institut Teknologi Kalimantan

\*Corresponding author e-mail : syamsul@lecturer.itk.ac.id

### ABSTRAK

*Youtube* merupakan *platform* video terbesar di dunia dengan total pengguna sebanyak 1,5 miliar pada tahun 2018. *Youtube* menjadi salah satu *platform* penyedia informasi, salah satunya yakni kenaikan harga minyak mentah dunia hingga berada di atas US\$100 per barel. Berdasarkan permasalahan tersebut, penulis melakukan penelitian terkait analisis sentimen dari komentar pengguna *Youtube* mengenai kenaikan harga BBM menggunakan metode *Gaussian naïve bayes*. Percobaan dilakukan menggunakan 3053 *dataset* dengan pelabelan menggunakan *lexicon* dan *split data* 8:2. Penerapan vektorisasi kata menggunakan *word embedding Fasttext* dan *Bag of word* sebagai pembandingan terhadap akurasi. Percobaan dilakukan dengan kombinasi perbedaan dimensi *size* pada proses pembuatan language model *fasttext*. Berdasarkan hasil penelitian yang telah dilakukan, didapatkan nilai akurasi tertinggi pada percobaan dengan *dataset* tanpa *filtering stopword* dan model *fasttext size* 100 dengan akurasi sebesar 74%. Berdasarkan hasil evaluasi, sistem yang dibangun dapat mengklasifikasikan sentimen atau opini publik ke dalam sentimen positif dan sentiment negatif secara otomatis.

**Kata kunci** : BBM, *Fasttext*, *Lexicon*, *Gaussian naïve bayes*, *Word embedding*

### ABSTRACT

*Youtube* is the largest video platform in the world with a total of 1.5 billion users in 2018. *Youtube* is one of the information provider platforms, one of which is the increase in world crude oil prices to above US\$100/barrel. Based on these problems, the authors conducted research related to sentiment analysis from *Youtube* user comments regarding the increase in fuel prices using the *Gaussian naïve Bayes* method. The experiment was carried out using 3053 datasets with labeling using *lexicon* and 8:2 data split. The vectorization uses *Fasttext* and *BoW* as a comparison of accuracy. The experiment was carried out with a combination of size dimensions *fasttext*. Based on the results of the research, the highest accuracy value was obtained in experiments with a dataset without *stopword* and *fasttext size* 100 with an accuracy of 74%. The system built can classify public sentiment into positive and negative sentiments automatically.

**Keywords**: Fuel, *Fasttext*, *Lexicon*, *Gaussian naïve bayes*, *Word embedding*

## I. PENDAHULUAN

*Youtube* merupakan *platform* video terbesar di dunia dengan total pengguna sebanyak 1,5 miliar pada tahun 2018 [1]. Dilansir dari hasil survei yang dilakukan oleh *Hootsuite (We Are Social)* mengenai *Indonesian Digital Report*, terdapat sebanyak 93,8% pengguna *Youtube* dari jumlah populasi di Indonesia pada tahun 2021. Peningkatan jumlah pengguna berbanding lurus dengan semakin bertambahnya konten yang diunggah di *Youtube*. *Youtube* dewasa ini

tidak hanya menjadi sarana hiburan bagi pengguna, namun juga sebagai sarana penyedia berbagai informasi terbaru. Oleh karena itu, dapat dikatakan bahwa *Youtube* merupakan *television modern* yang digemari oleh berbagai kalangan karena memuat konten yang bermacam-macam dan dapat dipilih sesuai keinginan pengguna.

*Youtube* juga menjadi sarana dalam menyampaikan pendapat atau opini dari setiap individu, baik melalui unggahan video maupun dari komentar pengguna. Pengguna dapat dengan bebas

memberikan opini melalui kolom komentar mengenai konten video yang ditonton. Melihat peningkatan pengguna *Youtube* di Indonesia serta kebebasan dalam memberikan opini pada komentar *Youtube*, hal ini dapat dibuat suatu pendekatan untuk mengetahui pendapat dari pengguna *Youtube* terhadap suatu konflik atau peristiwa besar yang sedang terjadi [2]. Salah satunya yakni mengenai Kenaikan Harga BBM yang merupakan salah satu dampak dari ketegangan yang terjadi antara Rusia dan Ukraina yang membuat harga minyak mentah dunia meroket hingga berada di atas US\$100 per barel [3].

Dilansir dari CNN Indonesia, kenaikan harga BBM diterapkan di seluruh daerah di Indonesia setelah dilakukan pertimbangan mengenai lonjakan dari harga minyak mentah Indonesia (ICP) yang awalnya tercatat pada bulan Desember 2021 sebesar US\$73,36 per barel menjadi US\$114,55 per barel pada 24 Maret 2022 [4]. Berdasarkan permasalahan tersebut, penulis melakukan penelitian terkait analisis sentimen dari komentar pengguna *Youtube* mengenai kenaikan harga BBM. Hal ini dilakukan untuk mengetahui tanggapan masyarakat terhadap keputusan pemerintah dalam menaikkan harga BBM di seluruh wilayah Indonesia. Banyaknya tanggapan pro dan kontra di masyarakat membuat penulis tertarik untuk mengangkat topik tersebut.

Pendekatan *text mining* dengan NLP (*Natural Language Processing*) dilakukan dalam menganalisa dan mencari makna dari setiap opini/pendapat pengguna *Youtube*. Analisa sentimen memerlukan metode dalam mengklasifikasikan teks, salah satunya yakni metode *Gaussian naïve bayes*. Metode *Gaussian naïve bayes* merupakan salah satu algoritma berbasis nilai kontinu dengan konsep probabilitas yang dapat digunakan dalam menentukan kelas dari dokumen dan mampu mengolah data dalam jumlah besar dengan akurasi yang tinggi [5]. Performa tingkat akurasi yang dihasilkan oleh metode *Gaussian naïve bayes* bergantung pada data latih yang dipilih, akurasi akan maksimal jika data latih mencakup sebagian besar atau bahkan mencakup keseluruhan dari data yang dimiliki [6]. Beberapa penelitian terdahulu yang dalam penerapannya menggunakan metode *Naïve Bayes Classifier* diantaranya yakni penelitian yang dilakukan oleh Ratnawati dkk [7], yakni menganalisa sentimen publik melalui Twitter dimana implementasi yang dilakukan memperoleh akurasi sebesar 90%, precision 92%, dan *recall* 90% [7]. Penelitian lainnya juga dilakukan oleh Herdhianto dkk [8], yang menghasilkan akurasi sebesar 74% pada penelitian menggunakan algoritma *naïve bayes* dengan ekstraksi fitur *term frequency*. Kekurangan yang terdapat pada kedua paper tersebut yakni, pada paper pertama tidak menggunakan ekstraksi fitur dalam penerapannya serta data latih yang digunakan terlalu sedikit akibatnya pada saat pengujian terdapat data yang

tidak dikenali sehingga terdapat perbedaan mengenai hasil prediksi dan label sebenarnya.

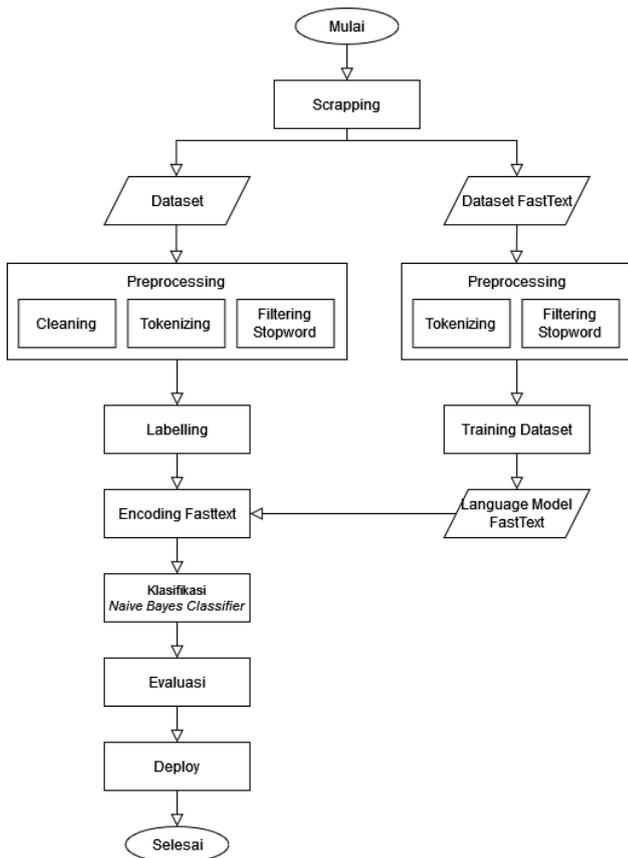
Ekstraksi fitur memiliki peran penting dalam melakukan klasifikasi teks dimana proses ini berfungsi mengubah kata menjadi nilai vektor sehingga sistem dapat menghitung probabilitas setiap komentar yang kemudian diklasifikasikan kedalam polaritas yang sesuai. Pada penelitian yang dilakukan oleh Ratniasih dkk [9], pelabelan dilakukan secara otomatis menggunakan *lexicon* dan hasil akurasi yang didapat kurang maksimal karena data yang digunakan terlalu sedikit. Penelitian lainnya dilakukan oleh Kusumawati dkk [10] mengenai analisa sentimen menggunakan ekstraksi fitur *fasttext* dimana pada penelitian tersebut memiliki kesimpulan bahwa vektorisasi kata merupakan tahap yang paling penting dalam melakukan klasifikasi teks. Parameter dimensi *size* pada language model *fasttext* memiliki pengaruh terhadap hasil performa yang didapat dimana pada penelitian tersebut menggunakan *corpus* dengan dimensi 300 sebagai model terbaik pada kasus yang diterapkan. *Fasttext* merupakan salah satu ekstraksi fitur yang berbasis *word vector* dengan memberikan nilai vektor pada setiap kata dan memperhitungkan keterkaitan antar kata. Salah satu kelebihan yang dimiliki oleh *Fasttext* yakni, mampu mengatasi kasus *out of vocabulary* (OOV) dimana *fasttext* mampu memberikan nilai vektor jika terdapat kata baru yang muncul.

Berdasarkan penelitian yang telah dilakukan tersebut, penulis tertarik melakukan penelitian menggunakan metode *Gaussian naïve bayes* pada analisis opini masyarakat mengenai kenaikan Harga BBM melalui komentar pengguna *Youtube* dengan ekstraksi fitur *Fasttext* untuk melihat pengaruh terhadap hasil akurasi yang didapatkan. Penelitian ini diharapkan dapat memperbaiki kekurangan-kekurangan yang terdapat pada penelitian terdahulu.

## II. METODE

### 1. Diagram Alir Penelitian

Tahapan dalam penelitian yang berjudul “Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar *Youtube* Dengan Metode *Gaussian naïve bayes*” dimulai dari proses pengambilan data (*scraping data*), pembersihan data (*preprocessing*), pelabelan data (*labelling*), pembuatan *language model fasttext*, *encoding fasttext*, klasifikasi *Gaussian naïve bayes*, evaluasi model dan *deployment*. Diagram alir dari penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram alir penelitian

## 2. Scraping

Tahap pertama yakni melakukan pengambilan data (*scraping*) dari salah satu video yang membahas mengenai kenaikan harga BBM pada platform Youtube. Pengumpulan data dilakukan dengan mengambil komentar pengguna youtube pada video yang membahas mengenai kenaikan harga BBM. Komentar tersebut yang nantinya akan dijadikan sebagai data dalam melakukan analisa sentimen masyarakat mengenai kebijakan pemerintah dalam menaikkan harga BBM.

## 3. Preprocessing

Tahap berikutnya adalah *preprocessing* berfungsi untuk membuang data yang tidak sempurna yang dapat menyebabkan gangguan pada data sehingga data tidak konsisten. Tujuan dilakukannya *preprocessing* agar data siap untuk digunakan dengan menghilangkan elemen yang dianggap tidak memiliki makna [11]. *Preprocessing* data dilakukan dengan beberapa tahapan yakni:

### 3.1 Cleaning

*Cleaning* merupakan tahap pembersihan data setelah proses pengambilan data dimana data masih mentah dan terdapat banyak simbol yang tidak dibutuhkan di dalamnya. Proses *cleaning* dilakukan dengan tujuan agar data bersih dari elemen yang dapat merusak model serta mempercepat proses komputasi. Proses ini dilakukan dengan menghilangkan kata-kata

diantaranya yakni url, username (@), hashtag (#), emoticon, kata yang dianggap spam, filter kata alay dan case folding.

### 3.2 Tokenizing

*Tokenizing* merupakan proses pemenggalan dari setiap kalimat yang ada menjadi potongan kata atau token. Proses ini dilakukan dengan tujuan untuk memudahkan dalam proses *stopword removal* dan *labelling* data agar lebih akurat.

### 3.3 Stopword Removal

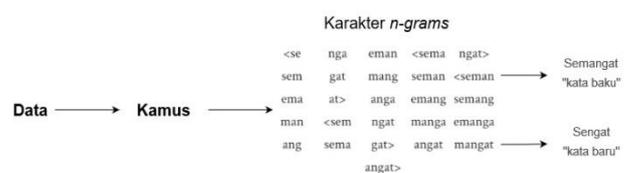
*Stopword Removal* merupakan tahap penghapusan kata yang terdapat pada *list stopwords* bahasa Indonesia. Proses *filtering stopwords* bertujuan untuk mengurangi kata yang umum dan kurang memiliki arti yang terdapat pada kamus *stopword*.

## 4. Labelling

Tahap *Labelling* merupakan tahap penentuan polaritas pada setiap komentar yang ada dengan memberikan bobot pada setiap kata yang terdapat di dalam dokumen [12]. Proses pelabelan data dilakukan secara otomatis menggunakan bantuan dari *Lexicon* agar proses pelabelan menjadi lebih cepat. Pelabelan dilakukan dengan memberikan bobot pada setiap kata kemudian menjumlahkan total pembobotan kata agar dapat ditentukan polaritas yang di dapat. Tujuan dilakukan pelabelan untuk melihat hasil dari pengelompokkan polaritas menggunakan *lexicon* yang nantinya dijadikan sebagai nilai actual [13].

## 5. Fasttext

Tahap selanjutnya yakni ekstraksi fitur menggunakan *Fasttext* yang merupakan salah satu metode vektorisasi kata dimana kata yang sudah bersih diberikan nilai numerik guna meningkatkan kinerja pada saat pengklasifikasian data [14]. Pembentukan vektor dilakukan dengan *sub-word embedding* yakni dengan memenggal setiap kata menjadi sub-kata dan mereduksi dimensi vektor menjadi lebih kecil. Vektor kata yang telah dilatih kemudian di encode untuk menjadi sentence vektor dengan menjumlahkan dan merata-ratakan nilai setiap vektor kata yang terbentuk. Ilustrasi *n-gram* pada *fasttext* dalam proses *sub-word embedding* ditunjukkan pada Gambar 2.



Gambar 2. Ilustrasi *n-gram* pada *fasttext*

Ilustrasi diambil dari kata “semangat” yang kemudian dipecah menjadi beberapa sub-kata (n-gram), kemudian pembentukan kata baru didapatkan berdasarkan hasil dari sub-kata pada kamus data yang digabungkan. Contohnya kata “sangat” yang merupakan hasil penggabungan dari sub-kata “se” dan “ngat”.

## 6. Klasifikasi Gaussian naïve bayes

Tahap klasifikasi dilakukan dengan menggunakan metode *Gaussian naïve bayes* yang merupakan metode pengklasifikasian data dengan menghitung probabilitas menggunakan Teorema Bayes. Tahap klasifikasi terbagi menjadi 2 tahap yakni tahap training dan testing terhadap model yang telah dibuat. Pembagian *dataset* dilakukan dengan metode split dimana data memiliki rasio 8:2 antara data training dan data testing. Pembagian *dataset* dilakukan secara random dari total keseluruhan data yang ada [2]. Data training digunakan untuk melatih model klasifikasi sedangkan data testing digunakan untuk melakukan evaluasi model [5]. Perhitungan probabilitas dalam metode *Gaussian naïve bayes* dilakukan dengan menggunakan rumus yang ditunjukkan pada Persamaan (1) dan Persamaan (2) [15].

$$p(x_i|y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \quad (1)$$

Keterangan:

$p(x_i|y)$  : Probabilitas  $x_i$  terhadap kelas  $y$

$x_i$  : Nilai Kelas- $i$

$\sigma$  : Nilai *Variance*

$\mu$  : *Mean* Kelas  $y$

$\pi$  : Phi 3.14

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2)$$

Keterangan:

X : merupakan data yang tidak diketahui kelasnya

H : hipotesis dari data X yang merupakan kelas spesifik

$P(H|X)$  : probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)

$P(H)$  : probabilitas hipotesis H (*prior probability*)

$P(X|H)$  : probabilitas X berdasarkan kondisi dari hipotesis H

$P(X)$  : probabilitas dari X

## 7. Evaluasi

Tahap Evaluasi merupakan tahap akhir sistem dalam proses analisis sentimen. Evaluasi dilakukan dengan tujuan untuk mengetahui hasil kinerja model yang sudah dilatih sebelumnya [6]. Tahap ini dilakukan dengan menggunakan *confusion matrix*

yang bertujuan untuk melakukan evaluasi terhadap klasifikasi yang sudah dilakukan dengan membandingkan hasil antara prediksi data yang didapat dengan data aktual. Setelah melakukan perbandingan terhadap kedua data, dilakukan perhitungan terhadap *accuracy*, *precision*, *recall*, dan *f1-score*.

## 8. Deployment

*Deployment* merupakan tahap akhir penelitian yang bertujuan untuk pengaplikasian sentimen analisis ke dalam sistem berbasis *website* dengan bantuan framework Flask. Pengaplikasian sistem bertujuan untuk dapat melakukan klasifikasi sentimen dengan cepat dan mudah yakni hanya dengan meng-input kalimat atau sentimen dan sistem akan menghitung probabilitas untuk menentukan polaritas sentimen apakah termasuk sentimen positif atau sentimen negatif.

## III. HASIL DAN PEMBAHASAN

### 1. Scraping

Tahap pertama yang dilakukan pada penelitian ini yakni *scraping* data atau pengambilan data sebagai bahan dalam melakukan penelitian mengenai analisis sentimen masyarakat. Data diambil dari komentar pengguna pada video *youtube* yang membahas mengenai kenaikan harga BBM di Indonesia. Proses *scraping* data ditunjukkan pada Tabel 1.

Tabel 1. Hasil *scraping data*

Index	Author	Text
1	pati wiring	Akhirnya semua kendaraan beralih ke pertalite dan terjadilah kekurangan stok pertalite terpaksa rakyat kecil beli pertamax
2	anak rekeh	Saya sebagai ojeg online sangat berat hati kecewa.
3	Steven Efendy96	Jangan risau maseehhh rejeki udah ada yg ngatur.... Ngopi maseehhh.... ðŸ™• ðŸ™•
---	...	...
3151	AmeCraft	Sangat miris gi mana ya nasib orang yang gk mampu
3152	Heru Setiawan	Mantap naik terus jangan nanggung....

Berdasarkan Tabel 1, terdapat kolom berupa *author* dan *text* dari komentar pengguna *youtube*. Data yang diperoleh dari 4 video berbeda dengan topik permasalahan yang sama mengenai kenaikan harga BBM di Indonesia yakni sebanyak 3152 komentar. Pengambilan data dilakukan dengan menggunakan *Youtube Comment Downloader* yang dikembangkan oleh Ahmed Shahriar Sakib.

### 2. Preprocessing

Tahap *preprocessing* dilakukan dengan beberapa tahapan yakni *cleaning* untuk

menghilangkan elemen yang tidak diperlukan serta membersihkan data dari simbol dan tanda baca lainnya. Selanjutnya *tokenizing*, tujuan dari proses ini adalah untuk melakukan pemenggalan setiap kata pada suatu kalimat/komentar. Terakhir yakni *stopword removal*, proses ini bertujuan untuk menghapus kata umum yang sering muncul dan terdapat pada *list corpus stopwords* Indonesia. Tujuan dari *stopword removal* adalah untuk mengurangi fitur atau kata pada saat vektorisasi kata. Proses *preprocessing* ditunjukkan pada Tabel 2.

Tabel 2. Hasil *preprocessing data*

<i>Text</i>	<i>Preprocessing</i>
Akhirnya semua kendaraan beralih ke pertalite dan terjadilah kekurangan stok pertalite terpaksa rakyat kecil beli pertamax	['kendaraan', 'beralih', 'pertalite', 'kekurangan', 'stok', 'pertalite', 'terpaksa', 'rakyat', 'beli']
Saya sebagai ojeg online sangat berat hati kecewa.	['ojeg', 'online', 'berat', 'hati', 'kecewa']
Jangan risau maseehhh rejeki udah ada yg ngatur.... Ngopi maseehhh.... δÿ• δÿ™•	['risau', 'maseehhh', 'rezeki', 'mengatur', 'ngopi', 'maseehhh']
...	...
Sangat miris gi mana ya nasib orang yang gk mampu	['miris', 'iya', 'nasib', 'orang']
Mantap naik terus jangan nanggung....	['mantap', 'nanggung']

Berdasarkan Tabel 2, data hasil *scraping* yang ditunjukkan pada kolom *text* kemudian dilakukan proses *preprocessing* dengan hasil seperti pada kolom *preprocessing*. Proses ini bertujuan untuk membersihkan data agar data menjadi bersih sehingga mempermudah proses labeling dan mempercepat proses komputasi.

### 3. Labelling

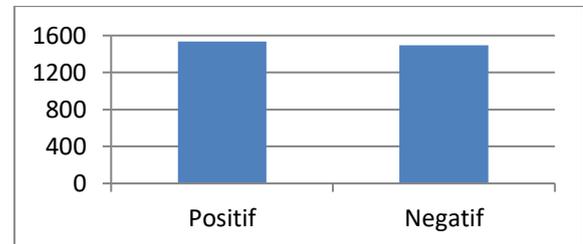
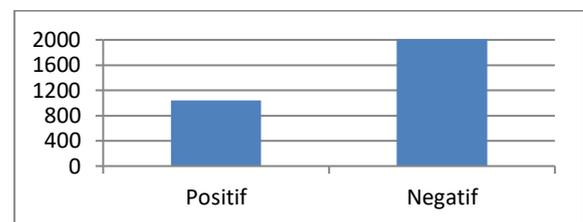
Proses selanjutnya yakni *labelling* data atau pemberian polaritas pada *dataset* yang sudah dibersihkan (*preprocessing*). Pengelompokan polaritas dilakukan secara otomatis menggunakan *Lexicon* dengan binary class atau 2 polaritas berbeda yakni positif yang bernilai 1 dan negatif yang bernilai -1. Proses *labelling lexicon* ditunjukkan pada Tabel 3.

Tabel 3. Hasil *labelling data*

<i>Preprocessing</i>	<i>Polarity Score</i>	<i>Polarity</i>
['kendaraan', 'beralih', 'pertalite', 'kekurangan', 'stok', 'pertalite', 'terpaksa', 'rakyat', 'beli']	-20	negatif
['ojeg', 'online', 'berat', 'hati', 'kecewa']	-8	negatif
['risau', 'maseehhh', 'rezeki', 'mengatur', 'ngopi', 'maseehhh']	0	positif
...	...	...
['miris', 'iya', 'nasib', 'orang']	-16	negatif
['mantap', 'nanggung']	3	positif

Berdasarkan Tabel 3, Pelabelan dilakukan dengan menjumlahkan setiap nilai pada angka sesuai

dengan bobot yang diberikan oleh kamus *lexicon*. Penentuan polaritas positif dan negatif berdasarkan total *polarity\_score* yang dijumlahkan, jika  $\geq 0$  maka diklasifikasikan sebagai label positif dan jika  $< 0$  maka diklasifikasikan sebagai label negatif. Hasil *labelling* data ditunjukkan pada Gambar 3 dan Gambar 4.

Gambar 3. Hasil *labelling dataset* dengan *stopword*Gambar 4. Hasil *labelling dataset* tanpa *stopword*

Kombinasi percobaan pada penelitian ini dilakukan dengan menggunakan *dataset* yang sama dengan perlakuan *preprocessing* data yang berbeda. *Dataset* pertama merupakan *dataset* tanpa *stopword* yang ditunjukkan pada Gambar 3 dan *dataset* kedua merupakan *dataset* dengan *stopword* yang ditunjukkan pada Gambar 4

### 4. Word embedding Fasttext

Tahap berikutnya yakni pembuatan language model *Fasttext* sebagai word vector dari *dataset* yang nantinya diklasifikasi. Pembuatan language model ini dilakukan untuk mencari nilai vektor dari setiap kata yang ada dengan memecah setiap kata kedalam sub-kata. Fitur *sub-word* dari *Fasttext* memungkinkan dalam mengatasi permasalahan mengenai *Out of vocabulary* (OOV), artinya *Fasttext* mampu mengatasi suatu kata yang tidak terbaca pada saat proses melatih data sehingga sistem mampu memberikan nilai vektor pada kata yang bahkan asing dan baru muncul sekalipun. Proses pemecahan kata menjadi n-gram pada *Fasttext* yang merupakan proses *sub-word embedding Fasttext* ditunjukkan pada Tabel 4.

Tabel 4. *Word vector fasttext*

<i>Index</i>	<i>Word Vector</i>
0	-0.064831
1	-0.492448
2	0.154022
3	-0.208632
4	0.071776
...	...

Berdasarkan Tabel 4, word *vector* yang didapat kemudian dilakukan proses encoding untuk merubah word *vector* menjadi sentence *vector*. Encoding dilakukan dengan membagi nilai vektor setiap kata terhadap jumlah kalimat kemudian dilakukan perhitungan rata-rata terhadap nilai vektor sehingga setiap kalimat terbentuk dari satu vektor utuh atau tabular.

**5. Klasifikasi Gaussian naïve bayes**

Tahap klasifikasi merupakan tahap perhitungan terhadap probabilitas setiap kata pada *dataset* terlabel dengan menggunakan metode *Gaussian naïve bayes*. *Dataset* yang telah dilakukan *labelling* kemudian di bagi menggunakan metode split data dengan rasio 8:2 yang terdiri 80% data latih dan 20% data uji. Data kemudian di latih dan diuji dengan menggunakan klasifikasi *Gaussian naïve bayes*. Hasil percobaan yang telah dilakukan terlampir di dalam Tabel 5.

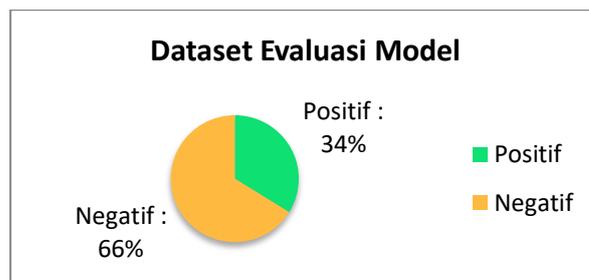
Tabel 5. Hasil percobaan klasifikasi

No	Dataset	Language model	Akurasi	
			Fasttext	Bag of word
1	Dataset dengan stopword	Fasttext size=100	60%	66%
		Fasttext size=200	60%	
		Fasttext size=300	59%	
2	Dataset tanpa stopword	Fasttext size=100	74%	68%
		Fasttext size=200	73%	
		Fasttext size=300	73%	

Berdasarkan Tabel 5, didapatkan hasil tertinggi pada saat penggunaan *dataset* tanpa *filtering stopword* dan menggunakan language model *fasttext* dengan dimensi 100, hasil akurasi yang didapat sebesar 0.74. Sedangkan hasil terendah didapat pada percobaan *dataset* dengan *filtering stopword* menggunakan language model *fasttext* dan dimensi 300 dengan akurasi sebesar 0.59.

**6. Evaluasi Model**

Evaluasi model merupakan tahap akhir dalam proses klasifikasi yang dilakukan dengan tujuan untuk menguji model yang telah dilatih dan diuji. Evaluasi model dilakukan dengan menggunakan data baru yang telah dilakukan *preprocessing* dan *labelling*. Data yang digunakan berjumlah 260 sentimen dengan perbandingan seperti pada Gambar 5.



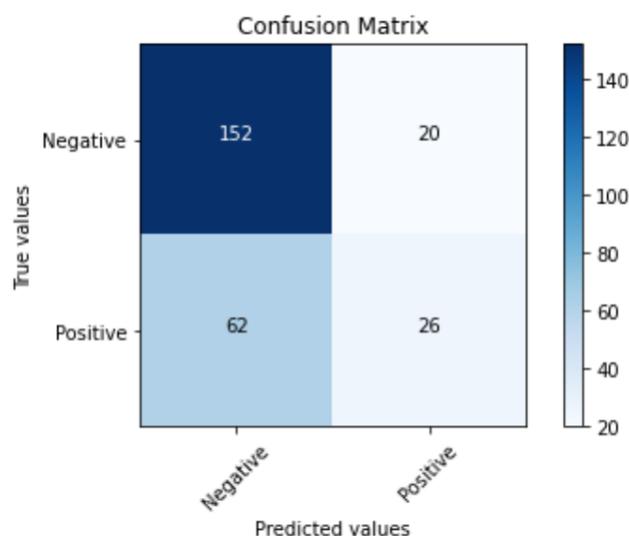
Gambar 5. Dataset evaluasi model

Total data evaluasi yang digunakan berjumlah 260 yang terdiri dari 172 sentimen negatif (-1) atau 66.2% dan 88 sentimen positif (1) atau 33.8%. Hasil evaluasi data ditampilkan dalam bentuk *confusion matrix* dengan menghitung hasil akurasi, presisi, *recall* dan *f1-score*. Hasil evaluasi model terlampir pada Gambar 6.

	precision	recall	f1-score	support
-1	0.71	0.88	0.79	172
1	0.57	0.30	0.39	88
accuracy			0.68	260
macro avg	0.64	0.59	0.59	260
weighted avg	0.66	0.68	0.65	260
True Negative :	152			
True Positive :	26			
False Negative :	62			
False Positive :	20			

Gambar 6. Nilai akurasi, presisi, *recall* dan *f1-score*

Berdasarkan Gambar 6, didapatkan nilai ilai akurasi sebesar 68%, nilai presisi sebesar 57%, *recall* sebesar 30% dan *f1-score* sebesar 39% dengan perbandingan antara data prediksi dan data aktual ditunjukkan pada Gambar 7.



Gambar 7. Visualisasi *confusion matrix*

Berdasarkan Gambar 7, model berhasil memprediksi 152 sentimen benar negatif dan 26 sentimen benar positif berdasarkan data aktual hasil dari pelabelan. Model berhasil memprediksi 62



1. Mengambil *dataset* dari video siaran langsung *youtube* agar mendapatkan data yang lebih bersih dari komentar antar pengguna *youtube*.
2. Menggunakan *corpus* atau kamus data *lexicon* dengan total vocab atau kata yang seimbang antara kata positif dan kata negatif.

## DAFTAR PUSTAKA

- [1] P. Y. Saputra, D. H. Subhi, and F. Z. A. Winatama, "IMPLEMENTASI SENTIMEN ANALISIS KOMENTAR CHANNEL VIDEO PELAYANAN PEMERINTAH DI YOUTUBE MENGGUNAKAN ALGORITMA NAÏVE BAYES," *J. Inform. Polinema*, vol. 5, no. 4, pp. 209–213, Aug. 2019, doi: 10.33795/JIP.V5I4.259.
- [2] F. H. Saputra and C. Cholifah, "Pengaruh Narasi Dalam Konten Vlog Channel Youtube 'Menjadi Manusia' Terhadap Sikap Dalam Menjaga Kesehatan Mental," *J. ISIP J. Ilmu Sos. dan Ilmu Polit.*, vol. 19, no. 1, pp. 11–22, May 2022, doi: 10.36451/J.ISIP.V19I1.163.
- [3] A. Erawati, M. Surif, and S. F. Dalimunthe, "Analisis Wacana Kritis Nourman Fairclough terhadap Jokowi yang Menyentil Menteriannya Mengenai Kenaikan Harga Minyak Goreng," *J. Pendidik. Tambusai*, vol. 6, no. 2, pp. 10653–10662, Jun. 2022, doi: 10.31004/JPTAM.V6I2.4114.
- [4] G. Hrp, N. A.-J. J. I. Komputer, undefined Ekonomi, and undefined 2022, "Analisis Dampak Kebijakan Perubahan Publik Harga BBM terhadap Perekonomian Rakyat Indonesia," *ummaspul.e-journal.id*, Available: <https://ummaspul.e-journal.id/JKM/article/download/3601/1261>.
- [5] A. Lestari, R. Perdana, ... M. F.-T. I. dan I., and undefined 2017, "Analisis sentimen tentang opini pilkada dki 2017 pada dokumen twitter berbahasa indonesia menggunakan naive bayes dan pembobotan emoji," *j-ptiik.ub.ac.id*, vol. 1, no. 12, pp. 1718–1724, 2017, Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/627>.
- [6] Y. A. Singgalan, "Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Systematic Literature Review," *J. Inf. Syst. Informatics*, vol. 3, no. 2, pp. 278–302, Jun. 2021, doi: 10.33557/JURNALISI.V3I2.125.
- [7] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, pp. 50–59, Jun. 2018, doi: 10.35314/ISI.V3I1.335.
- [8] A. Herdhianto, "Sentiment analysis menggunakan Naïve Bayes Classifier (NBC) PADA tweet tentang zakat," Jul. 2020, Available: <https://repository.uinjkt.ac.id/dspace/handle/123456789/53661>.
- [9] N. L. Ratniasih, N. Wayan, and N. Jayanti, "Sentiment Analysis of Stakeholder Satisfaction Measurement," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 16, no. 2, pp. 217–226, Apr. 2022, doi: 10.22146/IJCCS.72245.
- [10] N. D. Kusumawati, S. Al Faraby, and M. D. P., "Analisis Sentimen Mengenai Rencana Vaksinasi Covid-19 Menggunakan Support Vector Machine Dengan String Kernel," *eProceedings Eng.*, vol. 8, no. 5, Oct. 2021,
- [11] T. N. Fatyanosa and F. A. Bachtiar, "Classification method comparison on Indonesian social media sentiment analysis," *Proc. - 2017 Int. Conf. Sustain. Inf. Eng. Technol. SIET 2017*, vol. 2018-January, pp. 310–315, Feb. 2018, doi: 10.1109/SIET.2017.8304154.
- [12] R. Mahendrajaya, G. A. Buntoro, and M. B. Setyawan, "ANALISIS SENTIMEN PENGGUNA GOPAY MENGGUNAKAN METODE LEXICON BASED DAN SUPPORT VECTOR MACHINE," *KOMPUTEK*, vol. 3, no. 2, pp. 52–63, Oct. 2019, doi: 10.24269/JKT.V3I2.270.
- [13] G. B.-J. D. Informatika and undefined 2016, "Analisis sentimen hatespeech pada twitter dengan metode naive bayes classifier dan support vector machine," *researchgate.net*.
- [14] F. Alfariqi, ... W. M., and undefined 2020, "Klasifikasi Sentimen Pada Twitter Dalam Membantu Pemilihan Kandidat Karyawan Dengan Menggunakan Convolutional Neural Network Dan Fasttext Embeddings," ... *.telkomuniversity.ac.id*.
- [15] A. Abdurrahim, L. Syafa'ah, and M. Lestandy, "Sentiment analysis of Covid-19 vaccine tweets utilizing Naïve Bayes," *AIP Conf. Proc.*, vol. 2453, no. 1, p. 020052, Jul. 2022, doi: 10.1063/5.0094607.